An Information Theoretic Evaluation Criterion for 3D Reconstruction Algorithms

Rama Chellappa
University of Maryland,
College Park, MD 20742.
rama@cfar.umd.edu

Amit K. Roy Chowdhury
University of California
Riverside, CA 92521.
amitrc@ee.ucr.edu

Abstract

Even though numerous algorithms exist for estimating the structure of a scene from its video, the solutions obtained are often of unacceptable quality. To overcome some of the deficiencies, many application systems rely on processing more information than necessary with the hope that the redundancy will help improve the quality. This raises the question about how the accuracy of the solution is related to the amount of information processed by the algorithm. Can we define the accuracy of the solution precisely enough that we automatically recognize situations where the quality of the data is so bad that even a large number of additional observations will not yield the desired solution? This paper proposes an information theoretic criterion for evaluating the quality of a 3D reconstruction in terms of the statistics of the observed parameters (i.e. the image correspondences). The accuracy of the reconstruction is judged by considering the change in mutual information (or equivalently the conditional differential entropy) between a scene and its reconstructions and its effectiveness is shown through simulations. A brief discussion on the applicability of information theoretic criteria for other vision algorithms concludes the paper.

1 Introduction

Obtaining accurate 3D models from video using the structure from motion (SfM) approach [1], [2], is extremely important because of its diverse applications, ranging from multimedia to medical diagnosis. Yet the quality of many of the automatic 3D reconstructions leave much to be desired. This has led many researchers to analyze the sensitivity, robustness and statistical error characterization of the existing algorithms, trying to understand algorithm behavior and the characteristics of the natural phenomenon that is being modeled [3], [4], [5], [6], [7], [8], [9]. To overcome these errors, the tendency has been to add redundancy in the information processed. This raises the question as to how the redundant information affects the quality of the final solution. In this paper, we consider the situation where multiple reconstructions of the same scene are available (called intermediate or individual reconstructions, in this paper), that are combined together to obtain the final estimate (Figure (1)). We compute the incremental mutual information between the unknown 3D structure and increasing numbers of intermediate reconstructions.

Before proceeding to give a detailed description of the idea, we would like to draw the attention of the reader briefly to the area of model selection in statistics (AIC, BIC, MDL etc. [10]). The idea of fitting models to geometric data was formalized by Kanatani using a Geometric Information Criterion (GIC) [11]. However, a large number of SfM algorithms are not model based; they reconstruct individual point features of the scene. Our work tries to define the *quality of reconstruction* from point features in information theoretic terms. We also provide a discussion on the usefulness of information theoretic measures for evaluating computer vision algorithms.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **SEP 2004**	2. REPORT TYPE	3. DATES COVERED **00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE **An Information Theoretic Evaluation Criterion for 3D Reconstruction Algorithms**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Maryland,College Park,MD,20742**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Proceedings of the 2004 Performance Metrics for Intelligent Systems Workshop (PerMIS ?04), Gaithersburg, MD on August 24-26 2004

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

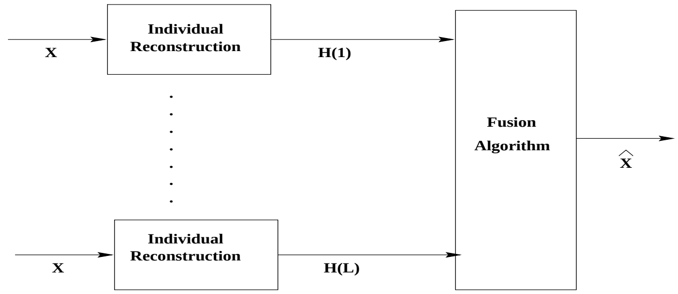

Figure 1: Block diagram representation of the reconstruction framework. $\mathbf{X}$ is the inverse depth that we want to estimate, $(\mathbf{H}(1), ..., \mathbf{H}(L))$ are the intermediate reconstructions (e.g. from each individual camera), and $\hat{\mathbf{X}}$ is the final fused estimate.

2 An Information Theoretic Criterion for 3D Reconstruction

2.1 Problem Formulation

We assume that all the depth values are aligned to a common frame of reference. Feature points will be represented by subscripts, separate reconstructions will be within parenthesis. The vector of estimates of the inverse depth [1] $[H_i(1), ..., H_i(N)]'$ will be denoted by $\mathbf{H}_i^{(N)}$. The boldface notation $\mathbf{H}(i)$ will represent all the features in the i^{th} reconstruction. The final estimate $\hat{\mathbf{X}}$ of $\mathbf{X} = [X_1, ..., X_M]'$ is obtained by fusing the individual reconstructions $(\mathbf{H}(1), ..., \mathbf{H}(L))$. To keep the notation simple, the subscript for the feature point will not be mentioned, unless required. The individual estimates are modeled as

$$H(i) = X + V(i) \tag{1}$$

where X is the inverse depth value of the particular feature.

2.2 Main Result

We will now present an information theoretic measure for evaluating the quality of a 3D reconstruction algortihm by analyzing the contribution of each of the individual reconstructions. Our entire analysis is for a particular point and thus the subscript will be dropped, unless required for clarity. Our criterion for evaluating the quality of reconstruction depends on estimating the difference in mutual information for the two sets of observations, $\mathbf{H}^{(L)}$ and $\mathbf{H}^{(L-1)}$. We term this as the *incremental mutual information* (IMI), i.e.

$$\Delta I(L) = I(X, \mathbf{H}^{(L)}) - I(X, \mathbf{H}^{(L-1)}). \tag{2}$$

The term gives us an idea of the contribution of the L^{th} observation to the reconstruction strategy with respect to the previous $(L-1)$ observations. As the number of observations increase, the effect of an additional observation decreases and approaches zero in the limit. In order to be assured that the reconstruction quality is actually improving, we need to consider only those situations where the mutual information $I(X, \mathbf{H}^{(L)})$ is non-decreasing. This ensures that we remove cases where the reconstruction is actually getting worse, and further observations are not improving it any more.

Using the relationship between mutual information and entropy, it is possible to obtain a different interpretation of the IMI. Denoting by $h(X)$ the entropy of the random variable X, we know that [12] $I(X;Y) = h(X) + h(Y) - h(X,Y)$. Thus $\Delta I(L)$ in (2) can be written as

$$\Delta I(L) \quad = \quad I(X; \mathbf{H}^{(L)}) - I(X; \mathbf{H}^{(L-1)}) \quad = \quad h(X|\mathbf{H}^{(L-1)}) - h(X|\mathbf{H}^{(L)}). \tag{3}$$

The quantity defined as the IMI can also be referred to as the incremental conditional entropy. Since entropy of a random variable is a measure of its uncertainty, ΔI measures the reduction in the uncertainty as we add an extra

[1] The inverse depth is used throughout this paper since it is the quantity that is estimated from the SfM equations for reconstruction from a video and its statistics can be obtained in an analytic form more easily than for the depth.

observation. Since the IMI tends to zero in the limit, the difference in the conditional entropy also approaches zero. Thus we will consider more and more images from the video sequence till the uncertainty in the final structure estimate can be reduced no further. This is the intuitive idea behind our criterion in (2).

The rate at which the IMI decreases is also an important measure of the progress of the algorithm. An extremely slow rate of fall indicates that more images will be necessary to achieve an acceptable level of quality. Since there is motion between adjacent frames of the video, a particular point will move out of the field of view of the camera after a certain amount of time. A very slow rate of fall of ΔI might mean that the quality of the reconstruction is not good enough even when the point is no longer visible. The rate of change of ΔI can be obtained as

$$\begin{aligned} \Delta^2 I(L) &= \Delta I(L) - \Delta I(L-1) \\ &= I(X, \mathbf{H}^{(L)}) + I(X, \mathbf{H}^{(L-2)}) - 2I(X, \mathbf{H}^{(L-1)}). \end{aligned} \tag{4}$$

Combining (2) and (4), we can state that an acceptable reconstruction quality has been achieved when $I(X, \mathbf{H}^{(L)})$ is non-decreasing **and** the following conditions are satisfied simultaneously:

$$\begin{aligned} \Delta^2 I(L) &\leq 0, \quad \forall L > L_0, \\ \Delta I(L) &< \tau, \end{aligned} \tag{5}$$

where L_0 is a constant and τ is a threshold defining an acceptable quality of reconstruction. Since $\Delta I(L)$ is monotone non-increasing for $L > L_0$ and is bounded below by zero, the monotone convergence theorem [13] applied to (3) implies that $h(X|\mathbf{H}^{(L-1)}) \to h(X|\mathbf{H}^{(L)}) \to h_0$ for some $L > L_0$. Thus, h_0 is the minimum level of uncertainty in a scene described by L observations.

Since the criterion does not depend on how the intermediate reconstructions are obtained, it is, in principle, independent of the 3D reconstruction strategy. However, the procedure for estimation of IMI may be optimized for a particular algorithm. Details on the estimation process can be found in [14].

2.3 IMI Computation Under Gaussian Distributions

Assume that $X \sim \mathcal{N}(0, \sigma_x^2 = P_X)$ and $\{V(i), i = 1, ..., N\}$ is a sequence of independent random variables distributed as $\mathcal{N}(0, \sigma_{V(i)}^2)$. Let $P_V = \text{diag}\,[P_V(i)]_{i=1,\dots,N} = \text{diag}\left[\sigma_{V(1)}^2, ..., \sigma_{V(N)}^2\right]$ [2].

From (1), $E[H(i)] = 0$ and

$$\begin{aligned} E[H(i)H(j)] &= E[(X + V(i))(X + V(j))] \\ &= P_X + P_V(i)\delta_{ij}, \end{aligned} \tag{6}$$

where δ_{ij} is a Kronecker delta function. Thus the covariance of $\mathbf{H}^{(N)}$ is $P_{\mathbf{H}^{(N)}} = P_V^{(N)} + \mathbf{1}_N P_X \mathbf{1}_N^T$, where $\mathbf{1}_N$ is a vector of N ones. Then the mutual information between X and $H(i)$,

$$\begin{aligned} I(X; H(i)) &= h(H(i)) - h(H(i)|X) \\ &= \frac{1}{2}\log\left(1 + \frac{P_X}{P_V(i)}\right). \end{aligned} \tag{7}$$

Next, consider the mutual information between the unknown X and the vector of observations $\mathbf{H}^{(N)}$. We will denote by $|K|$ the determinant of a matrix K.

$$\begin{aligned} I(X; \mathbf{H}^{(N)}) &= h(\mathbf{H}^{(N)}) - h(\mathbf{H}^{(N)}|X) \\ &\overset{(a)}{=} h(\mathbf{H}^{(N)}) - \sum_{i=1}^{N} \tfrac{1}{2}\log(2\pi e P_V(i)) \\ &\overset{(b)}{=} \tfrac{1}{2}\log\left(\frac{|P_V + \mathbf{1}_N P_X \mathbf{1}_N^T|}{|P_V|}\right). \end{aligned} \tag{8}$$

[2]Where necessary to distinguish a particular feature point, we will use the notation $\sigma_{x_j}^2$ and $P_{V_j}(i)$ or $\sigma_{V_j(i)}^2$ for the j^{th} point.

(a) is a result of applying the chain rule of entropy and substituting the expression for the differential entropy of a Gaussian random variable [12]; (b) is due to the fact that $|P_V| = \prod_{i=1}^{N} P_V(i) = \prod_{i=1}^{N} \sigma^2_{V(i)}$. Using the method of induction and the properties of determinants, it can be shown that $|P_V + \mathbf{1}_N P_X \mathbf{1}_N^T| = \prod_{i=1}^{N} \sigma^2_{V(i)} + \sigma_x^2 \sum_{i=1}^{N} \prod_{\substack{j=1 \\ j\neq i}}^{N} \sigma^2_{V(j)}$. Then from (8), the expression for the mutual information becomes

$$I(X;\mathbf{H}^{(N)}) = \tfrac{1}{2}\log\left(1 + \sum_{i=1}^{N} \frac{\sigma_X^2}{\sigma^2_{V(i)}}\right). \tag{9}$$

Let us compute the difference in the mutual information for the two sets of observations, $\mathbf{H}^{(N)}$ and $\mathbf{H}^{(N-1)}$. We shall call this the *incremental mutual information*, ΔI. Thus,

$$\begin{aligned}
\Delta I &= I(X;\mathbf{H}^{(N)}) - I(X;\mathbf{H}^{(N-1)}) \\
&= \tfrac{1}{2}\log\left(\frac{|P_{V^{(N)}} + \mathbf{1}_N P_X \mathbf{1}_N^T|}{|P_{V^{(N-1)}} + \mathbf{1}_{N-1} P_X \mathbf{1}_{N-1}^T|}.\frac{|P_{V^{(N-1)}}|}{|P_{V^{(N)}}|}\right) \\
&= \tfrac{1}{2}\log\left(\frac{\prod_{i=1}^{N} \sigma^2_{V(i)} + \sigma_x^2 \sum_{i=1}^{N} \prod_{\substack{j=1 \\ j\neq i}}^{N} \sigma^2_{V(j)}}{\prod_{i=1}^{N} \sigma^2_{V(i)} + \sigma_x^2 \sum_{i=1}^{N-1} \prod_{\substack{j=1 \\ j\neq i}}^{N} \sigma^2_{V(j)}}\right) \\
&= \tfrac{1}{2}\log\left(1 + \frac{1/\sigma^2_{V(N)}}{\frac{1}{\sigma_x^2} + \sum_{i=1}^{N-1} \frac{1}{\sigma^2_{V(i)}}}\right) \\
&= \tfrac{1}{2}\log\left(1 + \frac{1/P_V(N)}{\frac{1}{\sigma_x^2} + \sum_{i=1}^{N-1} \frac{1}{P_V(i)}}\right).
\end{aligned} \tag{10}$$

Equation (10) gives us a measure of the extra information that would be obtained by including an additional observation into the fusion process. Also, since

$$I(X;\mathbf{H}^{(N)}) - I(X;\mathbf{H}^{(N-1)}) = h(X|\mathbf{H}^{(N-1)}) - h(X|\mathbf{H}^{(N)}), \tag{11}$$

the quantity defined as the incremental mutual information can also be referred to as the incremental conditional entropy. Thus we are measuring the reduction in the uncertainty of the solution as we consider an extra observation. The difference in the differential entropy determines the decrease in the coding length of the scene structure as the number of observations increases [12].

The above calculation requires computing the variances of the intermediate reconstructions. Any method to compute them is perfectly suitable. In an earlier work [15], we have shown how to do this for the case of 3D reconstruction using optical flow. It should be remembered that all the geometric quantities have to be with respect to a particular frame of reference; hence it may be necessary to transform the variances appropriately.

An Estimation Theoretic Interpretation: We will now present an alternative interpretation of the result in (10) from an estimation theoretic perspective. The mean squared distortion is defined as

$$D(\mathbf{X},\hat{\mathbf{X}}) = \frac{1}{M}\sum_{j=1}^{M} E[(X_j - \hat{X}_j)^2]. \tag{12}$$

Let $p(X_j, H_j(1), ..., H_j(N))$ denote the joint density function of the parameter and observations. The mean square error estimator $\hat{X}_j$ of X_j, obtained from $\mathbf{H}^{(N)}$, is $\hat{X}_j(N) = E\left[X_j|H_j^{(N)}\right]$. From the Cramer-Rao lower bound

(CRLB) we can write the following set of inequalities.

$$
\begin{aligned}
D &\geq \frac{1}{M}\sum_{j=1}^{M}\frac{1}{E\left[-\frac{\partial^2}{\partial X^2}\log p(X_j, H_j(1),...,H_j(N))\right]} \\
&= \frac{1}{M}\sum_{j=1}^{M}\frac{1}{\frac{1}{\sigma^2_{x_j}} + \sum_{i=1}^{N} E\left[-\frac{\partial^2}{\partial X^2}\log p(H_j(i)|X)\right]} \\
&\geq \frac{1}{\frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{\sigma^2_{x_j}} + \sum_{i=1}^{N}\frac{1}{P_{V_j}(i)}\right)} \\
&\triangleq \frac{1}{\frac{1}{M}\sum_{j=1}^{M}\frac{1}{D_j(N)}}.
\end{aligned} \tag{13}
$$

The last step is a result of the application of Jensen's inequality [16] and the fact that $E\left[-\frac{\partial^2}{\partial X^2}\log p(H_j(i)|X)\right] = \frac{1}{P_{V_j}(i)}$. Recalling that (10) is for a particular feature point where the subscript has been suppressed for clarity of notation, let us denote $\Delta I_j \triangleq I(X_j;\mathbf{H}_j^{(N)}) - I(X_j;\mathbf{H}_j^{(N-1)})$. Then from (13) and the last expression of (10), we get

$$
\Delta I_j = \tfrac{1}{2}\log\left(\frac{D_j(N-1)}{D_j(N)}\right). \tag{14}
$$

Alternatively, the innovations at the N^{th} stage, $\gamma_N = X_N - \hat{X}_N$. Then following the standard derivation for the Kalman filter [16], it can be shown that variance of the innovations

$$
P_{\gamma_N} = \sigma^2_{V(N)}\left(1 + \frac{1/\sigma^2_{V(N)}}{\frac{1}{\sigma^2_x} + \sum_{i=1}^{N-1}\frac{1}{\sigma^2_{V(i)}}}\right), \tag{15}
$$

which shows that, for each feature point, the incremental mutual information is related to P_{γ_N} as

$$
\Delta I = \tfrac{1}{2}\log\left(\frac{P_{\gamma_N}}{\sigma^2_{V(N)}}\right). \tag{16}
$$

These relationships provide an alternative estimation theoretic interpretation to our result. Taken together (10), (14) and (16) demonstrate the use of statistical evaluation techniques to the SfM problem, when it is suitably formulated.

3 Analysis and Experiments

3.1 Analysis:

Present methods to evaluate the quality of a reconstruction involve computing the distortion in (12). For a fusion algorithm, this means that we need to compute (12) at every stage of the fusion and decide when to stop. This is computationally intensive, distortion measures are not always very useful in practical experiments since the choice of an acceptable threshold if often arbitrary and the source of the error (whether in the intermediate reconstructions or in the fusion algorithm) is difficult to identify. In our approach, (10) gives a direct way to measure the contribution of the intermediate solutions and the accuracy of the final solution as the algorithm progresses. The statistics of the error can be computed using the SfM equations and its solutions, as described in [15]. If the solution is far from its desired values, the error would be larger than if the solution is close to its true value. When the error in the intermediate reconstructions is small, D_j is small and hence the difference in the mutual information is small. Ideally, this difference should go to zero as we include more and more observations. If the error is large, D_j would be large and ΔI_j would not decrease appreciably with the number of observations. Another salient feature of our method is that we measure the information content between the true structure and the reconstructions *before* the

fusion. This allows us to understand the source of the error better since the effect of intermediate reconstructions and fusion algorithm are separated.

One scenario where this idea can be applied is reconstruction from a video sequence where intermediate reconstructions, $\mathbf{H}(1), ..., \mathbf{H}(L)$, obtained from a few frames (two or three) are combined together. Another application would be where partial reconstructions have been obtained from multiple cameras [3]. These partial models would have common overlapping regions which can be combined together to form the single estimate. In this case, $\mathbf{H}(1), ..., \mathbf{H}(L)$ would represent these common sub-regions from L separate reconstructions.

The statistical assumptions of independence and Gaussianity are necessary in order to derive closed form expressions for the quantities of interest. The independence of the intermediate estimates $H(1), ..., H(L)$ may be valid when these are obtained from separate imaging systems and then combined. When the same camera is used, the intermediate reconstructions should be obtained with non-overlapping frames; otherwise the common frames increase the dependencies. Regarding the Gaussianity assumptions, it has been pointed out by Zhang in [7] that the correspondence errors in SfM are usually normally distributed, if we can get rid of the outliers in the matches.

3.2 Experiments:

Experiment 1: A set of 3D points were generated so that we know their true positions. The perspective projections of these points were generated and Gaussian noise with zero mean and known variance was added to these 2D locations. The projections were taken for different positions of the camera, so that in the end a set of tracked features was obtained. From every pair of such tracked features, the positions of the original 3D points were estimated, which results in a set of 3D reconstructions. The first plot of Figure 2(a) shows the true value of the 3D points and their estimated reconstruction from all the frames over which the features could be tracked. [4] The second diagram in Figure 2(a) plots the decrease in the incremental mutual information with the increasing number of intermediate reconstructions.
Experiment 2: As in the previous simulation, a set of features were tracked over a number of frames. However, the level of noise added to the feature positions was higher and it led to a mismatch of some of the features. The 3D positions of the points were estimated using the SfM algorithm and the results were erroneous as is clear from the first plot of Figure 2(b). The second plot of Figure 2(b) depicts this case where the incremental mutual information remains large and does not follow any trend.
Experiment 3: We will now present our result on a real video sequence. The video consists of a person moving his head in front of a static camera. The aim was to reconstruct the model of the head of the person from this video. The focal length of the camera was known. Figure (3)(a) represents an image from the video along with some of the feature points which were tracked. Figure (3)(b) represents the change in the incremental mutual information between the unknown 3D structure and the intermediate reconstructions from every pair of frames. Based on this measure, the 3D model was reconstructed using 25 frames and Figure (3)(c) shows one particular view of this model.

4 A Discussion on the Usefulness of an Information Theoretic Criterion for Vision Algorithms

The statistical quality analysis of computer vision algorithms has been studied quite extensively (see [14] for a detailed literature survey on this topic). However, most of the methods have relied on computing the second order statistical moments, like covariance of the estimate. The covariance is a preferred measure because of its relation to the Cramer-Rao lower bound (CRLB), which dictates the minimum variance that an estimator can achieve [16]. If the variance of a sequence of estimates (say, of the 3D structure) tends towards the CRLB, then the estimate is said to be asymptotically efficient. However, computation of the CRLB often assumes that the estimate is unbiased (see [6]). This is because, computing the bias of an estimator is not an easy task. Hence, even though expressions exist

[3] This is the set-up in the "Eye Vision" technology developed by Carnegie Mellon University (CMU) and CBS Television (http://www.ri.cmu.edu/events/sb35/tksuperbowl.html).
[4] The first point was used to set the scale of the reconstruction, so that the geometric indeterminacies do not affect the result.

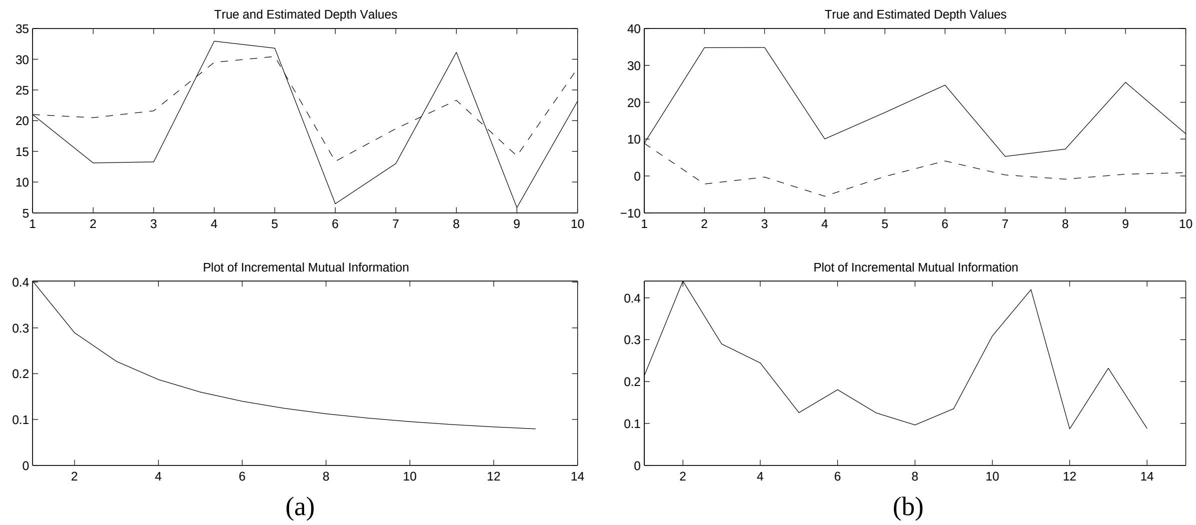

Figure 2: (a): The upper plot shows the true value of the depth of the 3D points using the solid line and the fused estimate from the intermediate reconstructions from all the frames using the dotted lines. The second diagram plots the decrease in the incremental information with the increasing number of frames. (b): The upper plot shows the true value of the depth of the 3D points using the solid line and the fused estimate from the intermediate reconstructions from all the frames using the dotted lines. The lower plot is the change in the mutual information with increasing number of frames. This is the case where the estimated reconstruction does not converge to the true value even with increasing observations.

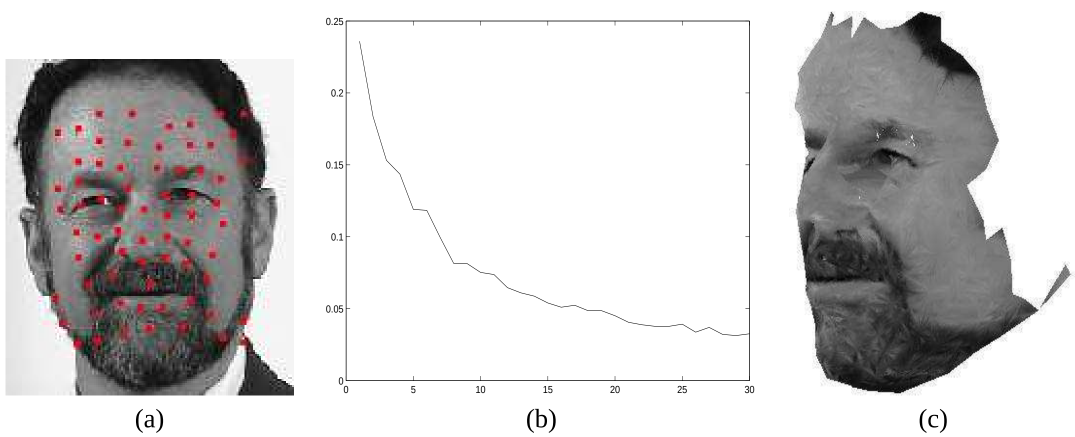

Figure 3: The above figures represent a 3D reconstruction from video using the method of measuring the IMI to judge the quality of the result. (a) is one of the images from the video along with the set of tracked features used for the reconstruction. (b) represents the change in the IMI with the number of images; (c) depicts one view from the reconstructed model.

for the CRLB of a biased estimator (known as the generalized CRLB), it is rarely used. The other main objection to the use of variance as a measure of quality is that it neglects the effect of higher order statistics. This is often a major approximation because the outliers, which are the source of many problems in computer vision algorithm, are often not modeled accurately by second order statistics.

Recent work [17, 18] has shown that the motion and depth estimates are statistically biased, and the bias is significant. This bias often propagates through later stages of the computation that rely on the motion and depth estimates. Also, as we have shown in [15], the noise in the SfM estimates is significantly non-Gaussian. Hence we propose that an information theoretic criterion which works by estimating the probability distribution function (pdf) of the concerned physical quantities (e.g. the depth), rather than concentrate on certain moments only, is a more suitable measure for a number of vision problems. The method of estimating the pdf will depend upon the particular algorithm and underlying assumptions. The major limitation of an information theoretic criterion is its efficient, robust and accurate estimation. This is because it is often difficult, and computationally expensive, to estimate the probability density functions of the parameters of interest. However, estimation of MI has received some attention

among researchers in signal processing and information theory [19]. It is our hope that such information theoretic criteria, as proposed in this paper, will become practically applicable as progress is made on robustly estimating them.

5 Conclusion

In this paper, we have introduced a method to evaluate the quality of 3D reconstruction from a video sequence. Existing methods rely on computing the distortion between the projections of the reconstructions and the original images and deciding that the reconstruction is of acceptable quality when the distortion is below a certain empirically chosen threshold. In this paper, we have shown that it is possible to evaluate the quality of the 3D structure estimate as the algorithm proceeds by computing the incremental mutual information, which determines the importance of considering an additional observation. It is related to the decrease in the coding length of the actual structure conditioned on the increasing number of observations. Finally, experimental results have been provided to justify these claims.

References

[1] O.D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, 1993.

[2] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[3] J. Oliensis, "A critique of structure from motion algorithms," *NECI Tech. Report*, vol. http://www.neci.nj.nec.com/homepages/oliensis/, 2000.

[4] Y. Ma, J. Kosecka, and S. Sastry, "Linear differential algorithm for motion recovery: A geometric approach," *International Journal of Computer Vision*, vol. 36, pp. 71–89, January 2000.

[5] J. Weng, N. Ahuja, and T.S. Huang, "Optimal motion and structure estimation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 864–884, September 1993.

[6] G.S. Young and R. Chellappa, "Statistical analysis of inherent ambiguities in recovering 3-d motion from a noisy flow field," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 995–1013, October 1992.

[7] Z.Y. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *International Journal of Computer Vision*, vol. 27, pp. 161–195, March 1998.

[8] D.D. Morris, K. Kanatani, and T. Kanade, "3d model accuracy and gauge fixing," in *CMU-RI-TR*, 2000.

[9] Z. Sun, V. Ramesh, and A.M. Tekalp, "Error characterization of the factorization method," *Computer Vision and Image Understanding*, vol. 82, pp. 110–137, May 2001.

[10] M. Hansen and B. Yu, "Model selection and the principle of minimum description length," *To appear in Journal of the American Statistical Association.*

[11] K. Kanatani, "Geometric information criterion for model selection," *IJCV*, vol. 26, no. 3, pp. 171–189, February 1998.

[12] T. Cover and J. Thomas, *Elements of Information Theory*, John Wiley and Sons, 1991.

[13] R. Walter, *Principles of Mathematical Analysis, 3rd Edition*, McGraw-Hill, 1976.

[14] A. Roy-Chowdhury and R. Chellappa, "An information theoretic criterion for evaluating the quality of 3D reconstructions," *IEEE Trans. on Image Processing*, In press.

[15] A. Roy-Chowdhury and R. Chellappa, "Stochastic approximation and rate-distortion analysis for robust structure and motion estimation," *Intl. Journal of Computer Vision*, vol. 55, pp. 27–53, October 2003.

[16] H.V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, 1988.

[17] C. Fermuller, D. Shulman, and Y. Aloimonos, "The statistics of optical flow," *Computer Vision and Image Understanding*, vol. 82, 2001.

[18] A. Roy Chowdhury and R. Chellappa, "Statistical error propagation in 3D modeling from monocular video," in *CVPR Workshop on Statistical Analysis in Computer Vision*, 2003.

[19] G.A. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space," *IEEE Trans. on Information Theory*, vol. 45, pp. 1315–1321, 1999.